

国立国語研究所学術情報リポジトリ

# 接続助詞とジャンル別文体的特徴の関連について： 『現代日本語書き言葉均衡コーパス』を資料として

著者	宮内 佐夜香
雑誌名	国立国語研究所論集
号	3
ページ	39-52
発行年	2012-05
URL	<a href="http://doi.org/10.15084/00000489">http://doi.org/10.15084/00000489</a>

## 接続助詞とジャンル別文体的特徴の関連について

——『現代日本語書き言葉均衡コーパス』を資料として——

宮内佐夜香

中京大学

国立国語研究所 コーパス開発センター 非常勤研究員 [–2011.03]

### 要旨

本稿は多様なジャンルを含む大規模コーパスを資料として、各ジャンルにおける接続助詞の出現傾向とジャンル別の文体的特徴との関連について論じたものである。分析の結果、接続助詞群の出現比率はジャンルごとに明らかに異なっており、その異なりはフォーマル、話し言葉的、丁寧等、ジャンルの文体的特徴との関連が認められ、個々の接続助詞が文体に関連する特徴を持つことが分かった。また接続助詞の出現傾向と文体的特徴の関連性を踏まえて、文体的特徴が未知であるジャンルについて、接続助詞群の出現比率を指標とした文体的特徴の検討を試みた\*。

**キーワード：**接続助詞、文体的特徴、ジャンル、大規模コーパス

### 1. はじめに

本稿は『現代日本語書き言葉均衡コーパス』(Balanced Corpus of Contemporary Written Japanese 以下 BCCWJ) に付与された形態論情報を利用して、複文を構成する接続助詞と、文体的特徴との関連を調査するものである。BCCWJ には、書籍、新聞、雑誌、白書、Web データ、教科書、法律等、多様なジャンル・媒体 (以下ジャンル) の書き言葉が収録されているが<sup>1</sup>、ジャンルによってその文体的特徴にはさまざまな差異が見られる。BCCWJ に基づいたジャンル間の文体的差異については、これまでに佐野・丸山 (2008)、小磯ほか (2008)、小磯・小木曾・小椋 (2008)、小磯ほか (2009) 等の諸研究が行われている。本稿ではこれらの先行研究によって明らかになった各ジャンルの文体的特徴や、ジャンル間の相違点、共通点を踏まえた上で、各ジャンルの接続助詞の出現比率を確認し、接続助詞の持つ文体に関連する特徴について考察する。また、接続助詞の出現傾向を手掛かりに、文体的特徴が未知であるジャンルの文体について検討することを試みる。

文体ごとの接続助詞使用の差異に関する研究はこれまでも行われてきたが、利用可能な大規模コーパスの不在により、対象とする資料が小規模にとどまることやジャンルに偏りが見られるなどの問題が生じている。大規模でありかつ多様なジャンルを収録する BCCWJ を対象とすることで、接続助詞の使用実態を数量的な面から明らかにすることができる。

\* 本稿は、大学共同利用機関法人人間文化研究機構国立国語研究所コーパス開発センターによる成果の一部を利用したものであり、内容は日本語学会 2009 年度春季大会、および特定領域研究「日本語コーパス」平成 21 年度全体会議での発表をもととしている。

<sup>1</sup> BCCWJ の概要は前川 (2008) 参照。

また大規模な電子化テキストデータがあっても、ある特定の語の調査を行う場合に、単純な文字列検索による抽出では多くの不要な用例が混在する可能性が大きい。特に本稿で扱う接続助詞類をはじめ語の長さの短い形式は容易には抽出できず、機械的な計量をするのが困難である。そのような短い形式でも、形態論情報が付与され品詞や活用形等が特定されたデータであれば、必要な用例を効率的に抽出することができる。BCCWJ に付与された形態論情報は、そうした側面においても有用であると考ええる。

## 2. 調査について

### 2.1 調査データ

本稿の調査は『現代日本語書き言葉均衡コーパス DVD 版 Ver. 1』に収録された形態論情報の表形式データ、及び、オンラインのコーパス検索アプリケーション「中納言」<sup>2</sup>を利用して行った。

BCCWJ に付与された形態論情報には、長単位・短単位という 2 種の言語単位がある。「国語研究所 | によって」と複合名詞や複合辞を一つの単位とするのが長単位、「国語 | 研究 | 所 | に | よっ | て」と短く切るのが短単位である。これら 2 種の語に対してそれぞれ品詞等の付加情報が付与されている<sup>3</sup>。本稿では短単位を語の単位として用いる。

分析対象としたのは、人手修正によって形態論情報を 99% の精度とした「BCCWJ コアデータ」のうちの、白書、新聞、書籍、Yahoo! 知恵袋（以下知恵袋）と、コアデータ以外の一部のサンプルに対して、コアデータ同様の人手修正を加えたデータである。後者は、コアデータ以外にも研究利用のための高精度データを増やすために、特に出版サブコーパスの書籍に含まれる小説のサンプルに対して人手修正を行ったもので、約 10 万語規模のデータである。いずれのデータも可変長データを用いた。コアデータ以外の高精度サンプルの ID リストは稿末に挙げる<sup>4</sup>。以上のデータをまとめて人手修正データと呼ぶ。

分析対象とするジャンル名と各ジャンルの人手修正データの語数の概要は表 1 の通りである。書籍は話し言葉的な要素がより多く含まれる小説類をはかと区別するために、サンプル ID に表示された図書館資料の日本十進分類法（NDC）の第 1 次区分の情報を利用して、〈文学以外〉と〈文学〉（NDC 9 類）を分けて扱う。

コアデータには出版サブコーパスの雑誌、特定目的サブコーパスの Yahoo! ブログも含まれるが、これらについてはサンプル内の内容が多岐にわたり文体にも影響が大きいと考えるため、本稿では除外した<sup>5</sup>。

<sup>2</sup> 中納言 RC2 (Released at 2011-11-10) URL: <https://chunagon.ninjal.ac.jp/>

<sup>3</sup> 形態論情報の詳細は小椋ほか（2011）参照。

<sup>4</sup> コアデータについては国立国語研究所コーパス開発センター（2011）参照。

<sup>5</sup> ブログの文体の混雑性については Maekawa et al.（2010）に言及がある。

表1 調査ジャンルと延べ語数（短単位数）

白書	約 20 万語
新聞	約 30 万語
書籍 1 〈文学以外〉	約 15 万語
書籍 2 〈文学〉	約 15 万語
知恵袋	約 10 万語

## 2.2 調査対象の形式

調査対象とする形式は、まず短単位の品詞情報において接続助詞となっているものである。さらに短単位において接続助詞以外の品詞情報が付与されている以下の形式も対象とした。条件表現に用いられる形式の「たら」（品詞情報は助動詞「た」の仮定形）「なら」（品詞情報は助動詞「だ」の仮定形）、また、短単位においては2語に解析される形式であるが、一般に接続助詞として認定される「ので」（準体助詞「の」+助動詞「だ」連用形）と「ものの」（名詞「もの」+格助詞「の」）も品詞情報を利用して抽出し、分析対象とした。本稿ではこれらの形式をまとめて接続助詞と呼ぶ。調査対象となるのは計13形式である。

このうち、「だから」「ですが」「なので」「ならば」「だったら」などの、機能語のみで構成される接続詞的連語の構成要素となっているものは分析対象外とした。また、接続助詞「て」は「ている」等の助動詞相当の連語や単純な文の接続において用いられ、ほかの形式と比較して極端に出現率が高くなるため、分析対象外とした。さらに全調査データ内の出現率が、総語数の0.005%未満の形式も分析対象外とした<sup>6</sup>。

短単位において2語以上に解析されているものの、接続表現に定型的に用いられる形式は、「ので」「ものの」のほかにも「のに」「ところで」「ために」「によって」等多数ある。これらには接続表現を担う形式として文法化している例のほか、実質的な意味を伴って用いられている例が高頻度で混在する。そのため現在付与されている形態論情報のみを用いて機械的处理で判別することが困難であり、目視確認による用法判別が必要になってくる。またその場合にも意味的な判別の困難な境界的な事例があり、どこまでを文法化した形式と考えるか、という問題も発生する。こうした多様性を持つ形式は本稿では分析対象外とした。大規模データにおいてこのような形式を調査するための方法については、今後の大きな課題であると考え<sup>7</sup>。

## 3. 接続助詞の出現傾向

### 3.1 文体的特徴に関する先行研究

まず、先行研究に基づいて各ジャンルの文体的特徴を確認する。樺島（1988）では、量的制限の強い文章（例えば新聞なら紙面の制限がある）ほど名詞比率が高くなることを指摘しており、

<sup>6</sup> 出現数が少ない形式の内訳は「けん」「さかい」「たって」「で」「ては」「ど」「とて」「なり」「に」（50音順）である。

<sup>7</sup> 例えば、機械学習による用法判別といった方向に展望があると考える。

最も名詞比率の低いものとして日常会話を挙げている。最も名詞比率が低いのが日常会話であることに着目すると、名詞比率は文章の持つ話し言葉らしさの程度を示す指標であるとも言えるかと思う。樺島（1988）から、本稿で扱うジャンルに関連する文章について名詞比率の増減を抜粋して示すと、新聞記事、小説地の文、小説会話の順に名詞比率が低くなっており、この順に話し言葉らしさの程度が高いと考えることができる。

次に小磯ほか（2008）、小磯・小木曾・小椋（2008）では、BCCWJのうち、白書、新聞、文学の文体の特徴について、名詞・動詞・機能語等の品詞ごとの出現率や、和語・漢語等の語種ごとの出現率などを指標として分析を行っている。これによると名詞率は新聞より白書の方が高くなっており、樺島（1988）の結果と合わせると、白書は新聞よりもさらに話し言葉らしさの程度が低いことが示唆される。また、小磯ほか（2008）、小磯・小木曾・小椋（2008）では機能語率にも着目して、「Halliday（1985）は、複雑な文章ほど動詞群の名詞化により機能語に対する内容語の比率が高くなるとし、内容語率で定義される語彙密度という尺度を提案した。《中略》機能語率の逆数が「内容語の占める割合」として語彙密度に概略相当すると考えるならば、文学、新聞、白書の順に文章としてより複雑であり、かつフォーマルということになる」（小磯ほか 2008: 193）と論じている。語彙密度については、佐野・丸山（2008）の分析がある。佐野・丸山（2008）では白書と文学を比較し、白書の方が語彙密度が高いという結果を得て、Halliday（1990）の「綿密に計画された、あるいはよりフォーマルな文章ほど語彙密度が高い」という主張との一致を指摘している。また佐野ほか（2009）においては、NDC 別に書籍の語彙密度を調査しており、書籍の中で、文学がもっとも語彙密度が低いという結果が示されている。

以上の BCCWJ を基とした先行研究の分析結果を、本稿で扱うジャンルに当てはめて考えると、書籍 2 〈文学〉、書籍 1 〈文学以外〉、新聞、白書の順に、よりフォーマルな特徴が見られることが分かる。この「フォーマルな」という概念は、「改まった」と言い換えることのできるものであると考える。「フォーマルでない」といった場合は「くだけた」特徴を指す。話し言葉であっても、改まったものとくだけたものがあり、書き言葉も同様である。「フォーマルさ」と「話し言葉的・書き言葉的」という概念は、このように異なる尺度であると言える。

さらに、小磯ほか（2009）では、書籍、新聞、白書に加え、知恵袋、国会会議録、『日本語話し言葉コーパス』（2005 年公開）内の学会講演と模擬講演（一般の人による個人的内容のスピーチ）の 4 ジャンルを取り上げ、計 7 ジャンルについて、小磯ほか（2008）と同様の分析を行っている。その結果「まず全体的な傾向として、追加した四つのジャンルはいずれも小説に近い傾向、つまり、漢語率、名詞率が低く、和語率、機能語率が高いという傾向を示している」（小磯ほか 2009: 595）と述べている。本稿で扱う知恵袋が、追加した他のジャンルと共に、小説と類似した傾向を持っていることが分かる。一方で、知恵袋の和語率が小説より低く、漢語率と名詞率が小説より高いという、白書等にやや近い数値であることを受けて、「質問という行為においてある程度改まった言葉遣いをすることや、回答・解説における専門性の高さなどが影響したものと考えられる」（小磯ほか 2009: 596）ともあり、知恵袋が一部小説と異なる傾向を持つことも指摘されている。

以上の先行研究から、書籍 2 〈文学〉、書籍 1 〈文学以外〉、新聞、白書の順に、よりフォーマルであること、また、この逆順に話し言葉的な特徴が強くなるという各ジャンルの文体的特徴を、本稿における前提として設定する。

以下、この前提に基づいて接続助詞の特徴について確認していく。ただし知恵袋については、Web 上の言語の研究は近年のものであるため先行研究が少ないことや、小説に近いというだけではなく白書等にも近い傾向があること（小磯ほか 2009）などを受けて、未知の部分の多いジャンルと考える。知恵袋については、文体的特徴が未知のジャンルとして、接続助詞の出現傾向を他のジャンルの出現傾向と比較することで、ジャンルの文体的特徴についての検討を試みることをとする。

### 3.2 接続助詞のジャンル別出現比率

各ジャンルの接続助詞の出現数を示したものが表 2、接続助詞の出現数の合計を 100 として、各形式の出現比率をジャンル別に示したものが表 3、出現比率をグラフで示したものが図 1 である<sup>8</sup>。表 2、表 3 には各形式の機能を示した。機能分類は『日本語機能表現辞書「つつじ」』に拠った<sup>9</sup>。

表 2 接続助詞のジャンル別出現数

機能分類	形式	白書	新聞	書籍 1 〈文学以外〉	書籍 2 〈文学〉	知恵袋	計
理由	から	37	112	308	327	246	1030
理由	ので	15	64	116	87	407	689
理由・添加	し	1	36	79	105	131	352
順接仮定	なら	3	54	78	103	165	403
順接確定	たら	1	51	154	175	291	672
順接仮定	ば	85	299	492	362	307	1545
順接仮定・順接確定	と	187	306	382	387	361	1623
逆接仮定	とも	6	14	12	16	3	51
逆接確定	が	195	564	567	450	890	2666
逆接確定	けれど	1	27	53	157	147	385
逆接確定	ものの	36	12	7	8	1	64
付帯	つつ	72	25	17	30	6	150
付帯	ながら	59	102	96	233	35	525
計		698	1666	2361	2440	2990	10155

<sup>8</sup> 同様の分析を自動解析の短単位データ延べ約 7000 万語分（形態素解析辞書 UniDic1.3.11 を用いて解析器「和布蕪（MeCab）」で解析）に対しても行ったところ、今回用いた人手修正データ延べ約 90 万語分に対する分析と同様の傾向が得られた。BCCWJ 全体の百分の一程度の規模のデータの傾向が規模の大きなデータの傾向とほぼ一致すること、また、人手修正を加えていない自動解析データの傾向が、高精度のデータと概ね一致することが、少なくとも今回の接続助詞に関する分析において確認された。

<sup>9</sup> 「つつじ」 URL: <http://kotoba.nuee.nagoya-u.ac.jp/tsutsuji> 松吉ほか（2007）参照。

表3 接続助詞のジャンル別出現比率

機能分類	形式	白書	新聞	書籍1 〈文学以外〉	書籍2 〈文学〉	知恵袋	全体の 比率
理由	から	5.3%	6.7%	13.0%	13.4%	8.2%	10.1%
理由	ので	2.1%	3.8%	4.9%	3.6%	13.6%	6.8%
理由・添加	し	0.1%	2.2%	3.3%	4.3%	4.4%	3.5%
順接仮定	なら	0.4%	3.2%	3.3%	4.2%	5.5%	4.0%
順接確定	たら	0.1%	3.1%	6.5%	7.2%	9.7%	6.6%
順接仮定	ば	12.2%	17.9%	20.8%	14.8%	10.3%	15.2%
順接仮定・順接確定	と	26.8%	18.4%	16.2%	15.9%	12.1%	16.0%
逆接仮定	とも	0.9%	0.8%	0.5%	0.7%	0.1%	0.5%
逆接確定	が	27.9%	33.9%	24.0%	18.4%	29.8%	26.3%
逆接確定	けれど	0.1%	1.6%	2.2%	6.4%	4.9%	3.8%
逆接確定	ものの	5.2%	0.7%	0.3%	0.3%	0.0%	0.6%
付帯	つつ	10.3%	1.5%	0.7%	1.2%	0.2%	1.5%
付帯	ながら	8.5%	6.1%	4.1%	9.5%	1.2%	5.2%
計		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

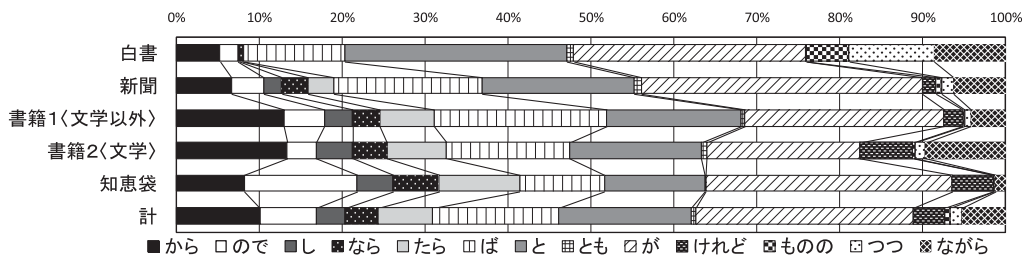


図1 接続助詞のジャンル別出現比率

表3, 図1を見ると, ジャンルによってはほとんど現れない場合のある形式には, 「し」「なら」「たら」「けれど」「ものの」「つつ」「ながら」などがある。これらはジャンル間の文体的特徴の差異に関わる特徴を持つ形式であると考えられる。はじめにこれらの, ジャンルによって出現比率に偏りのある形式から見ていこう。

白書において出現比率が特に低い形式として, 「けれど」「し」「なら」「たら」が挙げられる。それぞれの形式のジャンル別の出現比率を棒グラフにして並べたものを, 図2に示す。

未知のジャンルである知恵袋を除いて図2を見てみると, すべての形式の出現比率が白書で極端に低く, 新聞, 書籍1〈文学以外〉, 書籍2〈文学〉の順に出現比率が高くなるという同様の傾向が見られる。

代表として, 「けれど」のジャンル別出現比率を詳細に見ていく。表3を見ると白書が0.1%と

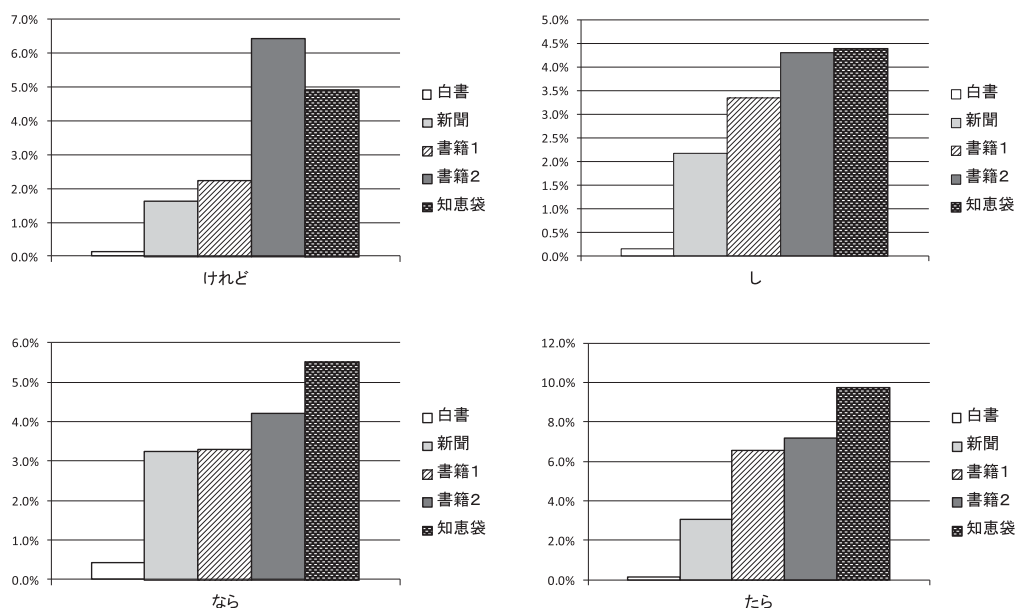


図2 「けれど」「し」「なら」「たら」のジャンル別出現比率

最も低く、書籍2〈文学〉が6.4%と最も高くなっている。ここから「けれど」はフォーマルではなく、かつ話し言葉的な特徴を持つ形式であると言える。「けれど」にはケレドとケドの異形態が含まれるが、その内訳を示すと表4のようになる<sup>10</sup>。

表4 ジャンル別「けれど」異形態用例数

	白書	新聞	書籍1 〈文学以外〉	書籍2 〈文学〉	知恵袋	計
ケレド	1	9	33	60	5	108
ケド	0	18	20	97	142	277
計	1	27	53	157	147	385

これを見ると「けれど」は多くケドの形態で現れており、書籍2〈文学〉においてケレドに対するケドの用例数が多いことが分かる。ガ・ケレド・ケド等の同じ機能を持つ形式のうち、ケドは最も話し言葉的であると考えられ<sup>11</sup>、話し言葉的な傾向の強い書籍2〈文学〉において「けれど」(ケド)の出現比率が高いことは頷ける結果である。

その他同様の傾向を持つ「なら」「たら」についても、話し言葉的な性質の形式であると言えるだろう。ただし、「けれど」と比較して、新聞や書籍1〈文学以外〉での出現比率が比較的高いことから、「けれど」(特にケド)に比べればややフォーマルな性質を持つ形式と言えそうである。「し」は「けれど」同様に新聞での出現比率が低いことから、「なら」「たら」よりも「けれど」

<sup>10</sup> ケレドモ、ケドモは短単位で「ケレド|モ」、「ケド|モ」と解析されるため、この中に含まれる。

<sup>11</sup> 永田・茂木(2007)では意識調査と談話資料をもとに、「ケド/ケレド/ケレドモ/ケドモ/ガ」の使い分けについて調査し、この中で「ケドが最も話し言葉的な特徴を持つ形式である」という結果を示している。



に近い、フォーマルでない性質がうかがえるのではないだろうか。

こうした傾向を踏まえて、知恵袋の文体的特徴を考えてみる。表3の数値を見ると、知恵袋では書籍2〈文学〉同様に、「けれど」の出現比率が4.9%と比較的高く、その形態のほとんどがケドである。また、「し」「なら」「たら」も書籍2〈文学〉同様出現比率が他のジャンルより高い。知恵袋の文章は、書籍2〈文学〉に含まれるような会話文とは異なりあくまで書き言葉であるが、書籍2〈文学〉と類似する傾向から考えて、話し言葉的な文体的特徴の強いジャンルであると指摘できる。

次に、白書に特に多く見られ、ほかのジャンルでは出現比率が低くなっている形式を見ていこう。「ものの」「つつ」がこれに当たり、「と」もそれに準ずる傾向が見られる。それぞれの形式ごとのジャンル別の出現比率を棒グラフにしたものを、図3に示す。

図3を見ると「ものの」と「つつ」は特に白書にのみ極端に多く使用され、他のジャンルでは出現比率が低い。「ものの」「つつ」は文体のフォーマルさや書き言葉の性質に関連する形式であると考えられるが、同じく書き言葉的である新聞や書籍1〈文学以外〉ではあまり用いられず、白書にのみ多く用いられることから、特に白書に特有のフォーマルさに関わる形式であるものと考えられる。

また、「と」は知恵袋以外の4ジャンル全体で「ものの」「つつ」ほどの大きな差はないが、白書、新聞、書籍1〈文学以外〉、書籍2〈文学〉の順に出現比率が低くなっていることが分かる。ここから、「と」は書き言葉的な性質に関連する形式であると思われる。これは同一の機能を持った「たら」が話し言葉的なジャンルに多用されているのと対照的な傾向であると言える。

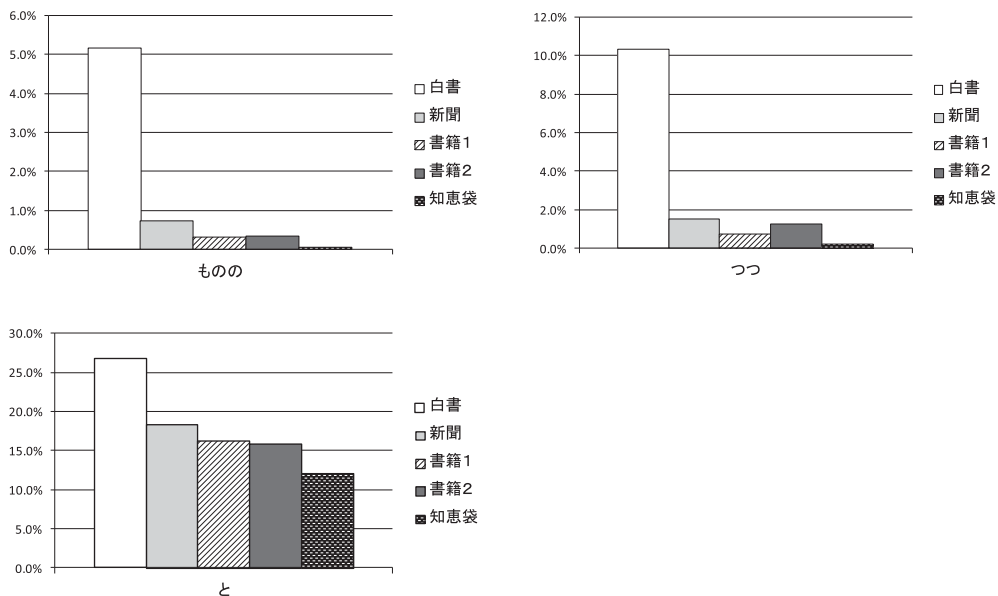


図3 「ものの」「つつ」「と」のジャンル別出現比率

知恵袋を見てみると、図3で分かるように知恵袋は「ものの」「つつ」両形ともほぼ用いられておらず、出現比率は「ものの」0.0%、「つつ」0.2%である（表3参照）。ここから、知恵袋は白書のフォーマルさと対極の性質を持っていると言えそうである。

出現傾向に偏りがあるものとして、最後に「ながら」「とも」について確認する。出現比率の棒グラフを図4に示した。

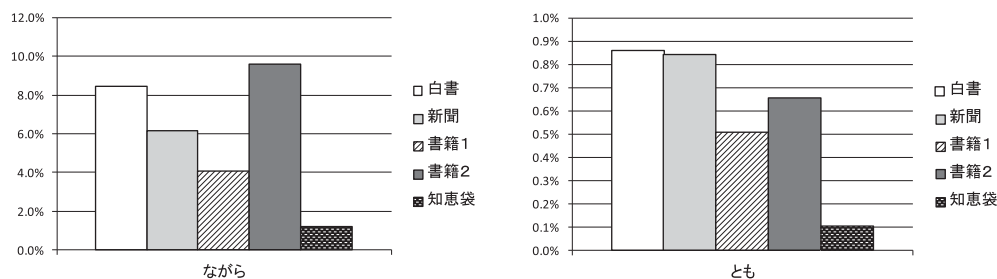


図4 「ながら」「とも」のジャンル別出現比率

表3の数値から、「ながら」の出現比率は、高い順に、書籍2〈文学〉9.5%、白書8.5%、新聞6.1%、書籍1〈文学以外〉4.1%、知恵袋が1.2%となっており、知恵袋で低い数値となっている。「とも」も同様の傾向が認められる。さきに話し言葉的である点で共通性を確認した書籍2〈文学〉と知恵袋が異なる傾向となっていることになる。

このことに着目し、書籍2〈文学〉と知恵袋の違いについて考えてみると、地の文の有無が挙げられる。書籍2〈文学〉のデータは文章の全体から取得したものであり、その中には小説などの会話文も、純粋に書き言葉である地の文も混在している。一方、知恵袋の文章には基本的に小説の会話文と地の文のような対立はなく、ケドの使用などから、知恵袋の文章は全体的に話し言葉的な性質を持っているものと考えられる。これにより「ながら」は書き言葉的な性質を持つ形式であることが示唆される。

しかし一方で「ながら」は、対立的な特徴を持つ白書と書籍2〈文学〉において出現比率がともに高いという傾向も持っており、「ながら」の特徴については今回前提としたフォーマルさ、話し言葉的・書き言葉的という特徴だけでは説明しきれない面もあると思われる。今後さらに検討すべきである。

出現比率に大きな偏りのない形式についても確認しておく。「が」「ば」がこれに当たる。出現比率の棒グラフを図5に示す。

どのジャンルにおいても概ね出現比率が高く、フォーマルさや話し言葉的・書き言葉的といった指標についてニュートラルな性質の形式であると考えられる。

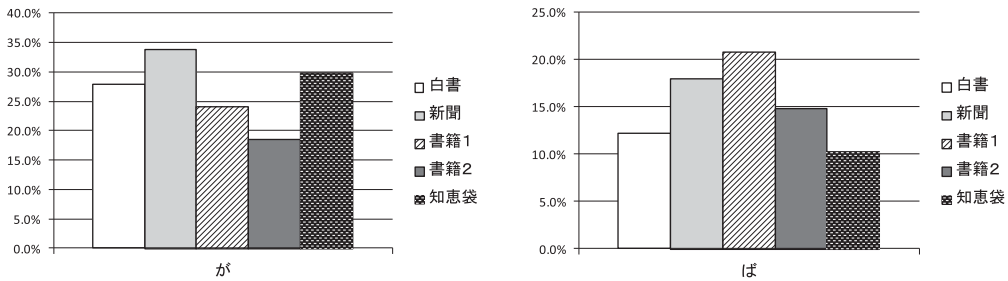


図5 「が」「ば」のジャンル別出現比率

### 3.3 機能別の出現傾向

次に、表2、表3に示した機能を参照し、同じ機能を持つ形式同士の出現傾向を比較して、文体的特徴について考える。ここでは特に理由表現について述べる<sup>12</sup>。

分析対象の接続助詞のうち、代表的な理由表現の形式として「から」「ので」が挙げられる<sup>13</sup>。出現比率の棒グラフを図6に、ジャンルごとの「から」「ので」のみの使用比率を図7に示した。

この「から」と「ので」はほとんどの資料で「から」の方の使用比率が高いが、知恵袋においては、「から」より「ので」の方の出現比率が高くなっている。そこで「から」と「ので」の差異について検討し、知恵袋特有の文体的特徴について考察してみたい。

「から」と「ので」の差異については、永野（1952）に詳しい。まず「から」が主観的で「ので」が客観的という差異が挙げられる。永野（1952）では、「から」は主節の述部に「推量・見解・意志・決心・命令・依頼・勧誘・質問など」の「話し手の主観に基づく表現」が現れ得るが、この場合「から」と「ので」は置き換えることができないことを指摘しており、「ので」は「客観的叙述」にのみ用いられると述べている。さらにその客観性故に「主観を押しつけない、淡々と述べている、という印象を与える」ために、「ので」が「丁寧な、やわらかい表現」であるとする。永野（1952）の指摘する「主観を押しつけない」「丁寧な、やわらかい表現」というのは対者的な待遇における性質であると言えるだろう。こうした「ので」の使用から、知恵袋は客観的な表現の多い文章であり、かつ対者的な配慮の見られる丁寧な文体的特徴を持っていることが示唆される。

知恵袋は質問掲示板であり、質問者は不特定のユーザーから必要な回答を得るために他者への配慮の見える丁寧な文体をとる傾向がある。また知恵袋には質問と質問者が選んだ「ベストアンサー」のセットが収録されているが、「ベストアンサー」は質問者の要望に的確に答えた詳細な内容であることが多く、また比較的丁寧な文体であることが多い。「ので」を含む知恵袋の文章の1例を（1）に示す。

<sup>12</sup> 逆接表現も「が」「けれど」「ものの」など複数の形式があるが、「が」「けれど」は「ものの」にはない逆接ではない接続の機能も担っている。「が」の多用にはその機能の広さも関連すると考えられ、文体との関連のみでは論じ難い。こうした機能別の考察についてはなお詳細な検討が必要である。

<sup>13</sup> 「し」については添加の意味も含むためここでは含めずに考える。

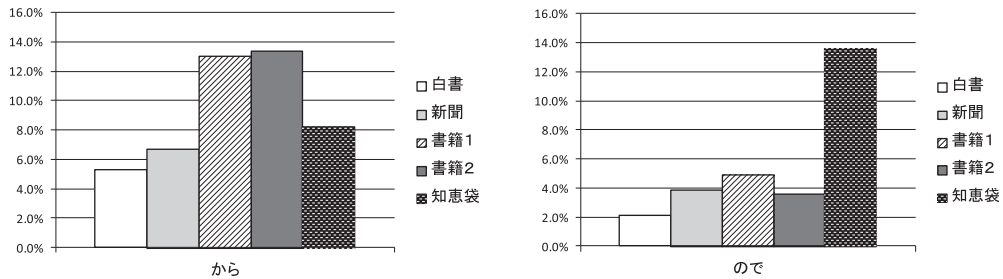


図6 「から」「ので」のジャンル別出現比率

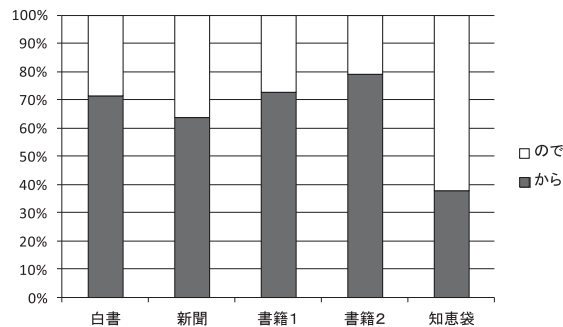


図7 「から」と「ので」の使用比率

## (1) [質問] メモリの増設を検討しています。

物理的にメモリを増設するほかに仮想メモリ？をCドライブ以外に作ることはできるのですか？

先ほどの回答に仮想メモリを1GB程作れば問題ないとの回答をもらいました。

HDDの容量が80GBあるのでそのほうがいいかな？と思ひまして。

あまり詳しくないので、勘違いしてたらすみません。

[回答] 先ほどのご質問が見当たらなかったのですが。。。。

物理メモリ（RAM）とHDD上に作成する仮想メモリではアクセススピードが仮想メモリのほうが段違いに遅いので、RAMが増設できるのであれば、RAMを増設したほうがよろしいかと思いますが。。。。

あ、ちなみに仮想メモリはCドライブ以外でも、作成はできます。

（知恵袋 カテゴリ：パソコン、周辺機器 サンプルID：OC02\_01040）

さらに質問の内容に対して詳細な回答が行われるために、知恵袋には専門的な解説が多く含まれていることから、客観的であるという性質もうかがえる。

ただし、同じく専門性が高いと思われる白書や新聞で「ので」がほとんど用いられていないという問題がある。この点については、「ので」は客観的であるという特徴以上に対者的な配慮という点で話し言葉的な性質が強く、書き言葉的な文体には現れにくいということが言えるのではないだろうか。

しかし、白書や新聞で「ので」が用いられないからといって、主観的な「から」が多用されるわけではない。表3に示した各ジャンルの「から」と「ので」の出現比率の数値を合計して比較すると、白書 7.4%、新聞 10.5%、書籍 1〈文学以外〉17.9%、書籍 2〈文学〉17.0%、知恵袋 21.8%となっており、そもそも白書と新聞では「から」「ので」がともにあまり用いられていないことが分かる。これは「から」と「ので」がともにフォーマルでなく、かつ書き言葉的な性質が弱いことを示唆している。その代わりに白書や新聞においては、理由表現形式としてどのような表現が用いられているかという点、本稿で扱っていない「ために」「によって」などの複合辞的な形式が多用されていることが考えられる。今後、複合辞についての詳細な調査が必要であるが、白書や新聞では話し言葉的な「ので」や主観的な「から」が用いられない代わりに、よりフォーマルかつ書き言葉的な他の形式が用いられているものと考えられる。

### 3.4 まとめ

以上、ジャンルごとの文体的特徴を踏まえて接続助詞の出現傾向を確認し、各形式の特徴について検討してきた。接続助詞がジャンルごとに異なる傾向で現れており、ジャンルの文体的特徴と関連した各形式の特徴が抽出できたものとする。

本稿で用いた文体的特徴に関するキーワードとしては、フォーマルさ、書き言葉的・話し言葉的、丁寧さ、客観的等が挙げられる。これまでの論に従ってそれぞれの接続助詞に対応するジャンルと特徴を示すと表5のようになる。

表5 接続助詞・ジャンル・文体的特徴の対応

接続助詞	ジャンル	文体的特徴
けれど・し・なら・たら	書籍 2〈文学〉・知恵袋	フォーマルでない話し言葉的
ものの・つつ	白書	フォーマル
と	白書・新聞・書籍 1〈文学以外〉	書き言葉的
ながら・とも	白書・書籍 2〈文学〉	書き言葉的
から・ので	書籍 1〈文学以外〉・書籍 2〈文学〉・知恵袋	フォーマルでない話し言葉的
ので	知恵袋	丁寧・客観的
が・ば	—	ニュートラル

最後に、前提として未知の部分の多いジャンルとしていた知恵袋の文体的特徴について、表5をもとにまとめると、知恵袋の文体には主に「フォーマルでない」「話し言葉的」「丁寧」「客観的」という特徴があることになる。

知恵袋の文体は、「けれど」「し」「なら」「たら」の出現傾向から、話し言葉的な性質が強いことが分かった。また、「ので」の多用は対者的な配慮に関連するものであり、知恵袋の文体の丁

寧さの現れと言えるだろう。対者的な配慮というのは直接的に読み手（聞き手）を想定しているということであり、この点も知恵袋の文体が話し言葉的であるということを示していると思われる。

3.1 で参照したように、小磯ほか（2009）は知恵袋の文体について、「質問という行為においてある程度改まった言葉遣いをすることや、回答・解説における専門性の高さ」の影響があり、白書等に若干近い性質があることを指摘している。この小磯ほか（2009）の指摘する「改まった言葉遣い」は、本稿で設定したフォーマルさに関連するようと思われるが、接続助詞の使用においては白書と共通するようなフォーマルな特徴は見出せなかった。接続助詞を指標として見ると、知恵袋には「ので」の使用で特徴付けられる客観的な記述が行われていると考えられる。この客観性は「回答・解説における専門性の高さ」に関連するものであると考えられ、白書等との共通点であると言える。この点においては小磯ほか（2009）の指摘との合致が見られた。

#### 4. おわりに

本稿では、BCCWJ に付与された形態論情報を用いて接続助詞の出現傾向について調査し、それを文体的特徴と関連付けて各形式の特徴について論じた。その結果、個々の接続助詞には明確に文体的特徴に関わる違いがあることが確認され、接続助詞の出現傾向が未知の文体的特徴について知るための指標となることが示唆された。

本稿では、短単位情報のみを利用して分析を行ったが、2.2 や 3.3 でも触れたように、接続表現を担う形式には今回対象としていない複合辞的な形式がある。こうした形式についても今後分析する必要がある。また、「ながら」や「し」、「が」など、複数の機能で用いられる接続助詞があるが、この用法の違いは BCCWJ に付与された品詞情報の範囲では判別されていない（この問題は複合辞的な形式の取り扱いと相通ずるものである）。このような複数の意味を持つ接続助詞について、機能別に傾向の違いがないかという問題も確認すべきものと思われる。さらに、本稿では分析対象外とした雑誌やブログ等の他の媒体についても、分析の必要があるだろう。その他、文体的特徴を示す尺度について、今回言及したもの以外も考えるべきである。これらについての調査・考察は今後の課題とする。

#### 書籍人手修正データリスト（コアデータ以外）

PB19_00035	PB29_00054	PB49_00027	PB59_00052
PB19_00036	PB29_00059	PB49_00030	PB59_00053
PB19_00043	PB29_00062	PB49_00041	PB59_00062
PB19_00047	PB29_00064	PB49_00049	PB59_00076
PB19_00073	PB29_00103	PB49_00067	PB59_00077
PB29_00022	PB39_00031	PB49_00075	PB59_00088
PB29_00034	PB39_00035	PB59_00031	
PB29_00044	PB39_00046	PB59_00033	
PB29_00053	PB39_00059	PB59_00047	

## 参考文献

- Halliday, M. A. K. (1985) *Spoken and written language*. Victoria: Deakin University.
- Halliday, M. A. K. (1990) Some grammatical problems in scientific English. *Annual Review of Applied Linguistics* 6: 13–37.
- 樺島忠夫 (1988) 『日本語はどう変わるか—語彙と文字—』東京：岩波書店.
- 小磯花絵・小木曾智信・小椋秀樹 (2008) 「短単位情報に基づくジャンル間の文体に関する分析」『特定領域研究「日本語コーパス」平成 20 年度全体会議予稿集』99–106.
- 小磯花絵・小木曾智信・小椋秀樹・富士池優美・宮内佐夜香 (2008) 「『現代日本語書き言葉均衡コーパス』にもとづくジャンル間の文体差に関わる要因の分析」『社会言語科学会第 22 回大会発表論文集』192–195.
- 小磯花絵・小木曾智信・小椋秀樹・宮内佐夜香 (2009) 「コーパスに基づく多様なジャンルの文体比較—短単位情報に着目して—」『言語処理学会第 15 回年次大会発表論文集』594–597.
- 国立国語研究所コーパス開発センター (2011) 「『現代日本語書き言葉均衡コーパス』利用の手引 第 1.0 版」『現代日本語書き言葉均衡コーパス DVD 版 Ver. 1』Disk 1.
- 前川喜久雄 (2008) 「KOTONOHA『現代日本語書き言葉均衡コーパス』の開発」『日本語の研究』4(1): 82–95.
- Mackawa, Kikuo, Makoto Yamazaki, Takehiko Maruyama, Masaya Yamaguchi, Hideki Ogura, Wakako Kashino, Toshinobu Ogiso, Hanae Koiso and Yasuharu Den (2010) Design, compilation, and preliminary analyses of Balanced Corpus of Contemporary Written Japanese. *Proceedings of LREC2010, Malta*: 1483–1486.
- 松吉俊・佐藤理史・宇津呂武仁 (2007) 「日本語機能表現辞書の編纂」『自然言語処理』4(5): 123–146.
- 永野賢 (1952) 「「から」と「ので」とはどう違うか」『国語と国文学』29(2): 31–41.
- 永田良太・茂木俊伸 (2007) 「接続助詞のスタイルをどう捉えるか—母語話者の意識調査とコーパスの分析から—」『語文と教育』21(1)–(8).
- 小椋秀樹・小磯花絵・富士池優美・宮内佐夜香・小西光・原裕 (2011) 『『現代日本語書き言葉均衡コーパス』形態論情報規程集 第 4 版 (上) (下)』国立国語研究所内部報告書.
- 佐野大樹・丸山岳彦 (2008) 「システミック文法に基づく書きことばの複雑さ測定—日本語大規模コーパスを用いた語彙密度計測—」『言語処理学会第 14 回年次大会予稿集』: 1097–1100.
- 佐野大樹・丸山岳彦・山崎誠・柏野和佳子・秋元祐哉・稲益佐知子・田中弥生・大矢内夢子 (2009) 『語彙密度を利用した『現代日本語書き言葉均衡コーパス』テキスト分類の試み』国立国語研究所内部報告書.

## The Relationship between Conjunctive Particles and Genre-related Style in the BCCWJ

MIYAUCHI Sayaka

Chukyo University

Adjunct Researcher, Center for Corpus Development,  
National Institute for Japanese Language and Linguistics [–2011.03]

### Abstract

This study examines how styles affect the frequency of uses of conjunctive particles in each style using a large-scale corpus including various genres. The results showed that the proportion of each conjunctive particle differs clearly in each genre, and that it represents the style-related characteristics of each genre such as formal, colloquial, polite, and so on. Moreover, as a result of examining the conjunctive particles in a genre whose style-related characteristics is unknown, it was suggested that the proportion of each conjunctive particle could be used as an index to know the style-related characteristics of a genre.

**Key words:** conjunctive particles, genre-related style, large-scale corpus